LATENT TOPIC MODELING OF WORD CO-OCCURRENCE INFORMATION FOR SPOKEN DOCUMENT RETRIEVAL

Berlin Chen

Department of Computer Science & Information Engineering,
National Taiwan Normal University, Taipei, Taiwan
berlin@csie.ntnu.edu.tw

ABSTRACT

In this paper, we present a word topic model (WTM) approach, discovering the co-occurrence relationship between words as well as the long-span latent topic information, for spoken document retrieval (SDR). A given document as a whole is modeled as a composite WTM model for generating an observed query. The underlying characteristics and different kinds of model structures are extensively investigated, while the performance of WTM is thoroughly analyzed and verified by comparison with a few existing retrieval models on the TDT-2 SDR task. We also attempt to incorporate part-of-speech (POS) weighting into the representations of the query observations and the WTM models for obtaining better retrieval performance.

Index Terms **—**language model, spoken document retrieval, word topic model, probabilistic latent semantic analysis

1. INTRODUCTION

Statistical language modeling (LM), which aims to capture the regularity in human natural language and quantify the acceptability of a given word sequence, has continuously been a focus of active research for a wide variety of speech recognition and natural language processing tasks over the past three decades. This statistical paradigm was first introduced for building information retrieval (IR) systems in the late 1990s [1, 2], indicating very good potential, and was then extended in a number of following publications [3, 4]. In general, these approaches attempt to build a probabilistic language model explicitly for each individual document in the collection, while the basic idea is that a document is deemed to be relevant to a query if the corresponding document model is more likely to generate the query.

In practice, the relevance measure for the LM approaches is usually computed by two different matching strategies, namely, literal term matching and concept matching [5]. The *n*-gram-based [1] and Hidden Markov model (HMM)-based [2, 4] approaches are the most popular examples for literal term matching. In these approaches, each document is interpreted as a generative model composed of a mixture of *n*-gram probability distributions for observing a query, while the query is considered as observations, expressed as a sequence of indexing terms (or words). However, most of these approaches would suffer from the problem of word usage diversity (or so-called vocabulary mismatch), which will make the retrieval performance degrade severely as a given query and its relevant documents are using quite a different set of words. In contrast, concept matching tries to explore the latent topic information conveyed in the query and documents, based on which the retrieval is performed; the probabilistic latent semantic analysis (PLSA) [3] is often considered as a representative of this category. PLSA introduces a set of latent topic variables to describe the "word-document" co-occurrence characteristics. The relevance measure of a query and a document is not computed directly based on the frequency of the query words occurring in the document, but instead based on the frequency of these words in the latent topics as well as the likelihood that the document generates the respective topics, which in fact exhibits some sort of concept matching. PLSA is usually trained in an unsupervised way [3] by maximizing the total log-likelihood of the document collection.

Over the last few years, spoken document retrieval (SDR) has attracted a lot of research attention from the speech processing community. In particular, the key aim has been focused on developing robust indexing (or representation) techniques so as to extract probable spoken terms or phrases inherent in a spoken document that could match the query words or phrases literally (the so-called spoken term detection, STD), instead of revolving around the notion of relevance of a spoken document to a query, through the use of existing retrieval models [6]. Nevertheless, a document is relevant if it could address the stated information need of the query, not because it just happens to contain all the words in the query [7].

In this paper, we propose a word topic model (WTM) approach to discover the occurrence dependence between words, as well as the long-span latent topic information, for the SDR task. Each document of a collection is modeled as a composite WTM model for predicting an observed query. The underlying characteristics and different kinds of model structures are investigated, while the performance of WTM is analyzed and compared with a few popular retrieval models.

2. LATENT TOPIC MODELING APPROACHES

2.1. Probabilistic Generative Framework

In information retrieval (IR), the relevance measure between a query Q and a document D can be expressed as $P(D|Q)$, i.e., the probability that the document D is relevant given that the query Q was posed, which can be transformed to the following equation by applying Bayes' rule:

$$P(D|Q)=\frac{P(Q|D)P(D)}{P(Q)}, \tag{1}$$

where $P(Q|D)$ is the probability of the query Q being generated by the document D, $P(D)$ is the prior probability of document D being relevant, and $P(Q)$ is the prior probability of query Q being posed. $P(Q)$ in Eq. (1) can be eliminated because it is identical for all documents and will not affect the ranking of the documents. Furthermore, because the way to estimate the probability $P(D)$ is still under active study [1, 2, 4], we may simply assume that $P(D)$ is uniformly distributed, or identical for all documents. In this way, the documents can be ranked by means of the probability $P(Q|D)$ instead of using the probability $P(D|Q)$. If the query Q is treated as a sequence of input observations (words or terms), $Q=w_1w_2...w_N$, where the query words are assumed to be conditionally independent given the document D (the so-called "bag-of-words" assumption), the relevance measure $P(Q|D)$ can be decomposed as a product of the probabilities of the query words generated by the document:

$$P(Q|D)=\prod_{w_i\in Q}P(w_i|D)^{c(w_i,Q)}, \tag{2}$$

where $c(w_i,Q)$ is the number of times that a specific word w_i occurring in Q.

2.2. Probabilistic Latent Semantic Analysis (PLSA)

In the PLSA modeling approach for IR, each individual document D can be interpreted as a document topic model (DTM), denoted as M_D, in which a set of K latent topics characterized with unigram (or multinomial) distributions are used to predict the query terms, and each of the latent topics is associated with a document-specific weight. That is, each document D can belong to many topics and the probability of a query word w_i generated by D is expressed by

$$P_{\text{PLSA}}(w_i|M_D) = \sum_{k=1}^{K} P(w_i|T_k)P(T_k|M_D), \quad (3)$$

where $P(w_i|T_k)$ denotes the probability of a certain type of query word w_i occurring in a specific latent topic T_k, and $P(T_k|M_D)$ is the posterior probability (or weight) of topic T_k conditioned on the document model M_D, with the constraint $\sum_{k=1}^{K} P(T_k|M_D)=1$ imposed. More precisely, the topic unigram distributions, e.g. $P(w_i|T_k)$, are shared among the entire DTM models, while each model M_D has its own probability distribution over the latent topics, e.g. $P(T_k|M_D)$. The key idea we wish to illustrate here is that the relevance measure of a query word w_i and a document D is not computed directly based on the frequency of w_i occurring in D, but instead based on the frequency of w_i in the latent topic T_k as well as the likelihood that D generates the respective topic T_k, which in fact exhibits some sort of concept matching [5]. The likelihood of a query Q generated by D is thus represented by

$$P_{\text{PLSA}}(Q|M_D) = \prod_{w_i \in Q} P_{\text{PLSA}}(w_i|M_D)^{c(w_i,Q)}. \quad (4)$$

In the practical implementation of PLSA [3], the corresponding DTM models are usually trained in an unsupervised way by maximizing the total log-likelihood of the document collection $\mathbf{D}$ in terms of the unigram $P_{\text{PLSA}}(w_i|M_D)$ of all words w_i observed in the document collection, or more specifically, the total log-likelihood of all documents generated by their own DTM models, using the Expectation-Maximization (EM) training algorithm [7]:

$$\begin{aligned} \log L_{\mathbf{D}} &= \sum_{D\in\mathbf{D}} \log P_{\text{PLSA}}(D|M_D) \\ &= \sum_{D\in\mathbf{D}} \sum_{w_i\in D} c(w_i, D) \log P_{\text{PLSA}}(w_i|M_D). \end{aligned} \quad (5)$$

2.3. Word Topic Model (WTM)

In this paper, we exploit an alternative probabilistic latent topic approach for information retrieval. Instead of treating each document in the collection as a document topic mixture model, we regard each word w_j of the language as a word topic model (WTM) M_{w_j} for predicting the occurrences of a particular word w_i:

$$P_{\text{WTM}}(w_i \mid M_{w_j}) = \sum_{k=1}^{K} P(w_i \mid T_k)P(T_k \mid M_{w_j}), \quad (6)$$

where $P(w_i \mid T_k)$ and $P(T_k \mid M_{w_j})$ are respectively the probability of a certain type of word w_i occurring in a specific latent topic T_k and the probability of the topic T_k conditioned on M_{w_j}. During the retrieval process, we can linearly combine the associated WTM models of the words involved in a document D to form a composite WTM model for D, and the likelihood of a query Q being generated by D can be expressed by

$$P_{\text{WTM}}(Q|M_D) = \prod_{w_i\in Q}\left[\sum_{w_j\in D} P_{\text{WTM}}(w_i|M_{w_j})P(w_j|D)\right]^{c(w_i,Q)}, \quad (7)$$

where $P(w_j \mid D)$ is estimated according to the relative frequency of w_j in D. In this way, the relevance measure between a query and document is determined by the product of a weighted sum of the probabilities that the respective WTM models of the words involved in the document generating each query word, and the documents having the highest probabilities expressed by Eq. (7) are therefore believed to be more relevant to Q.

The WTM models can also be optimized by the EM algorithm either with or without supervision. For unsupervised training of WTM, each WTM model M_{w_j} can be trained by concatenating those words occurring within a vicinity of, or a word context window of size S (S was experimentally set to 21 in this study) around, each occurrence of w_j, which are postulated to be relevant to w_j, in the spoken document collection to form a relevant observation sequence O_{w_j} for training M_{w_j}. The words in O_{w_j} are also assumed to be conditionally independent given w_j. Therefore, the WTM models of the words in the vocabulary $\mathbf{w}$ can be estimated by maximizing the total log-likelihood of their corresponding relevant observation sequences respectively generated by themselves, using the EM algorithm:

$$\begin{aligned} \log L_{\mathbf{w}} &= \sum_{w_j\in\mathbf{w}} \log P_{\text{WTM}}(O_{w_j}|M_{w_j}) \\ &= \sum_{w_j\in\mathbf{w}} \sum_{w_i\in O_{w_j}} c(w_i, O_{w_j}) \log P_{\text{WTM}}(w_i|M_{w_j}). \end{aligned} \quad (8)$$

In addition to unsupervised training of WTM, we also investigate supervised training of WTM in this paper. That is, given a training set of query exemplars and the associated query-document relevance information, the WTM models can be optimized by instead finding the model parameters that can maximize the total log-likelihood of the training set of query exemplars $\mathbf{Q}_{TrainSet}$ generated by their relevant documents:

$$\log L_{\mathbf{Q}_{TrainSet}} = \sum_{Q\in\mathbf{Q}_{TrainSet}} \sum_{D\in\mathbf{D}_{R\text{ to }Q}} \log P_{\text{WTM}}(Q|M_D), \quad (9)$$

where $\mathbf{D}_{R\text{ to }Q}$ denotes the set of documents that are relevant to a specific training query exemplar Q. Such a training approach in essence has the ability to associate documents with a query exemplar even though they do not share any of the query words. A similar treatment of supervised model training can also be applied to PLSA.

It is noteworthy that in recent years, the use of training query exemplars and the respective query-document relevance information (or the click-through information that to some extent reflects users' relative preferences of document relevance) also has been extensively studied for training various machine-learning based retrieval models like SVM (Support Vector Machines) [7, 8].

2.4. Theoretical Analysis of WTM and PLSA

WTM and PLSA can be analyzed from several perspectives. First, for unsupervised model training, PLSA models the co-occurrence relationship between words and documents, while WTM models the co-occurrence relationship between words, which is achieved by discovering the vicinity (or surrounding context) information of all occurrences of each vocabulary word in the document collection.

Second, the topic mixture weights $P(T_k|M_D)$ of PLSA for a new document D have to be estimated online using EM training, no matter whether the training is conducted in a supervised or unsupervised manner; on the contrary, for unsupervised training, the topic mixture weights $P(T_k|M_D)$ of WTM for a new document D can be simply estimated on the basis of the topic mixture weights $P(T_k|M_{w_i})$ of words w_j involved in the document without using the time-consuming EM training procedure.

Third, for unsupervised EM training, PLSA has $V\times K+K\times M$ parameters (cf. Eq. (3)) to be estimated and WTM has $V\times K\times 2$ parameters (cf. Eq. (6)); V denotes the size of the vocabulary set, while M the number of the documents and K the number of the latent topics. It is obvious that the number of the parameters of WTM needing unsupervised EM estimation will be larger than that of PLSA, when the number of training documents is less than the number of distinct words in the vocabulary ($N<V$). The number of the parameters of PLSA to be estimated grows linearly with the number of documents used for training PLSA, whereas that of WTM

instead remains the same regardless of the number of training documents used for unsupervised EM training, as the IR system adopt a closed set of vocabulary. Recently, a latent Dirichlet allocation (LDA) method [9] has been developed to address the above issue for PLSA. However, such a method still requires an iterative variational inference procedure for online estimating the associated parameters of a newly observed document.

Finally, it should be noted that for unsupervised training, if the context window for modeling the vicinity information of WTM is reduced to one word ($S=1$), WTM can be either degenerated to a unigram model as the latent topic number K is set to 1, or viewed as analogous to a bigram model (as $K=V$) or an aggregate Markov model (as $1<K<V$) [10]. Thus, with the appropriate values of S and K being chosen, WTM seems to be a good way to approximate the bigram or skip-bigram models for sparse data. One the other hand, WTM can be regarded as close in spirit to the word class-based model (WCBM) as well, by relating the latent topics of the former to the word classes of the latter [11]. WTM differs from WCBM in that WTM disregards word order information and leverages word co-occurrence statistics from longer text spans, whereas most of the approaches to using WCBM are based purely on modeling word bigram sequences.

3. EXPERIMENTAL SETUP

We used the Topic Detection and Tracking collection (TDT-2) for this work. The Chinese news stories (text) from Xinhua News Agency were used as our queries (or query exemplars) [4]. The Mandarin news stories from Voice of America news broadcasts were used as the spoken documents. All news stories were exhaustively tagged with event-based topic labels, which served as the relevance judgments for performance evaluation. Table 1 describes the details for the TDT-2 collection. The average word error rate obtained for the spoken documents is about 35%. The retrieval results, assuming manual transcriptions for the spoken documents to be retrieved (denoted TD, text documents) are known, are also shown for reference, compared to the results when only the erroneous transcriptions by speech recognition are available (denoted SD, spoken documents). The retrieval results are expressed in terms of non-interpolated mean average precision (MAP) [7].

In this paper, when PLSA and WTM are employed in evaluating the relevance between a query word w_i and a document D , we additionally incorporate the unigram probabilities of w_i in the document $P(w_i \mid D)$ and a general text corpus $P(w_i \mid Corpus)$ into PLSA and WTM, respectively, for probability smoothing and better performance. For example, the probability of a query word w_i generated by the WTM model of a word w_j (i.e., $P_{\text{WTM}}\left(w_i \mid \text{M}_{w_j}\right)$ in Eq. (6)) is therefore modified as follows:

$$\hat{P}_{\text{WTM}}\left(w_i \mid \text{M}_{w_j}\right) = \left(1-\rho_1-\rho_2\right)\cdot P_{\text{WTM}}\left(w_i \mid \text{M}_{w_j}\right) + \rho_1 \cdot P\left(w_i \mid D\right) + \rho_2 \cdot P\left(w_i \mid Corpus\right), \quad (10)$$

where ρ_1 and ρ_2 are weighting parameters ($0<\rho_1,\rho_2<1$ and $\rho_1+\rho_2<1$). Similar treatments also have been studied for PLSA [3].

4. EXPERIMENTAL RESULTS

4.1. Results on WTM

The first set of experiments aims at evaluating the retrieval performance of the word topic models trained with supervision (denoted as WTM-S) and varying model complexities. The model parameters were trained using a set of 819 training query exemplars (different from the test queries) with their corresponding query-document relevance information to the TDT-2 collection [4]. It should be borne in mind that, from now on, unless otherwise stated, the retrieval results reported were obtained by evaluating the ranked

Table 1: Statistics of the TDT-2 collection.

No. of spoken documents	2,265 stories, 46 hrs of audio		
No. of distinct text test queries	16 Xinhua text stories (Topics 20001~20096)		
	Min.	Max.	Mean
Document length (characters)	23	4841	287
Query length (characters)	183	2623	533
No. of relevant documents per query	2	95	29

list of documents returned by the retrieval models in response to each of the test queries. The retrieval results of WTM-S are shown in the upper part of Table 2, where each column illustrates the retrieval results in both the TD and SD cases by using different numbers of latent topics for modeling WTM-S. As can be seen, the retrieval performance is steadily improved as the topic number increases. The best retrieval result of 0.7672 is obtained for the TD case when the topic number is set to 128, while the best result is 0.7558 for the SD case with the same topic mixture number. Notice that although the word error rate (WER) for the spoken document collection is higher than 35%, the average degradation in retrieval performance is much smaller, especial when the topic mixture number becomes larger. Such an observation indicates that the WER does not cause much adverse effect on retrieval performance, which is quite in parallel with those reported by other groups [6]. One possible reason is that, a specific word (or phrase) might occur repeatedly (more than once) within a broadcast news story and it is not always the case that all the occurrences of the word would be misrecognized totally as other words.

In practical situations, the retrieval systems, however, might not have query exemplars correctly labeled with the query-document relevance information to be utilized for model training. Thus, we also study unsupervised model training for WTM (denoted as WTM-U). As evident in the lower part of Table 2, the performance of WTM-U is not always improved as the topic number increases. The best result of 0.6395 for the TD case is obtained when the document topic number is set to 128, while the best result of 0.5739 for the SD case when document topic number is 32. When comparing with the best results achieved in supervised training, there are at most about 0.13 and 0.18 decreases in the MAP measure, respectively, for the TD and SD cases.

To recap, for WTM, given a training set of query exemplars with the corresponding query-document relevance information, the retrieval results obtained based on the supervised training approach (WTM-S) are much better than those based on the unsupervised approach (WTM-U). Our hope is that, given a set of real user queries and the associated click-through information about the retrieved relevant documents, the performance of retrieval systems might be incrementally improved through use [4, 8].

4.2. Comparison of WTM and PLSA

Conventionally, PLSA is trained in a purely unsupervised manner. The retrieval results of such a modeling approach (denoted as PLSA-U) are shown in the lower part of Table 3. The best retrieval result of 0.6277 is obtained for the TD case when the latent topic number is set to 2, while the best result is 0.5681 for the SD case with 8 topic mixtures. They are slightly inferior to those achieved by WTM trained without supervision (WTM-U), but are markedly worse than those achieved by WTM trained with supervision (WTM-S).

We also explore supervised training for PLSA (denoted as PLSA-S), as described in Section 2.3. The same set of training query exemplars are employed here again to estimate parameters of PLSA.

Table 2: Retrieval results achieved by WTM.

No. Latent Topic		2	8	32	128
WTM-S	TD	0.6505	0.6887	0.7351	0.7672
	SD	0.5731	0.6186	0.6864	0.7558
WTM-U	TD	0.6336	0.6359	0.6382	0.6395
	SD	0.5693	0.5734	0.5739	0.5737

Table 3: Retrieval results achieved by PLSA.

No. Latent Topic		2	8	32	128
PLSA-S	TD	0.6362	0.6750	0.6823	0.7243
	SD	0.5759	0.5918	0.6255	0.6652
PLSA-U	TD	0.6277	0.6266	0.5949	0.6041
	SD	0.5545	0.5681	0.5534	0.5484

Table 4: Retrieval results achieved by VSM, LSA, HMM, and SVM, respectively.

Retrieval Model	VSM	LSA	HMM/ Unigram	HMM/ Bigram	SVM
TD	0.5548	0.5510	0.6327	0.5427	0.5797
SD	0.5122	0.5310	0.5658	0.4803	0.5317

Table 5: Retrieval results achieved by WTM with POS weighting.

No. Latent Topic		2	8	32	128
WTM-S	TD	0.6723	0.7007	0.7506	0.7858
	SD	0.5958	0.6331	0.6896	0.7554
WTM-U	TD	0.6512	0.6539	0.6542	0.6534
	SD	0.5774	0.5785	0.5807	0.5828

The corresponding retrieval results are shown in the upper part of Table 3, in which the best result of 0.7243 is obtained for the TD case when the document topic number is set to 128 and the best result of 0.6652 is obtained for the SD case with the same topic number. Such results are better than those obtained by using either WTM or PLSA trained in an unsupervised manner (WTM-U or PLSA-U), but are considerably worse than those obtained by using the WTM trained in a supervised manner (WTM-S). We can thus conclude that for the SDR task studied here, WTM is truly a good alternative to PLSA when the retrieval models are trained either with or without supervision.

4.3. Comparison of WTM and Other Retrieval Models

Moreover, we also compare WTM with other popular retrieval models, including vector space model (VSM), latent semantic analysis (LSA) [7], HMM, and SVM. The retrieval results of these four models are listed in Table 4 for comparison. VSM and LSA are implemented with the best parameter settings, while HMM and SVM are trained with supervision using the same set of 819 training query exemplars. Both the unigram and bigram modeling strategies are investigated for HMM [4]; on the other hand, the input to SVM consists of eleven commonly-used features, including the six features employed in [12]. As can be seen, WTM significantly outperforms all these four retrieval models when supervised learning is adopted (WTM-S). Even though WTM is trained in an unsupervised manner (WTM-U), its retrieval performance is still apparently better than that of VSM, LSA and SVM, and achieves quite competitive results to that of the HMM trained in a supervised manner. It is interesting that the performance of HMM degrades as the model structure becomes more sophisticated (i.e., from unigram to bigram modeling), whereas the performance of WTM and PLSA tends to becomes better as the topic number increased, when both models were trained with supervision. In brief, the LM approaches (WTM, PLSA and HMM) seems to be superior to SVM for the SDR task studied here; similar results were also observed in [12] for text document retrieval.

4.4. Incorporation of POS Weighting into WTM

Since the semantics of the proper nouns (such as locations, organizations, personal names, etc.), nouns and verbs are easier to identify and grasp than the adjectives, adverbs and connectives, they are expected to play more salient roles in the retrieval tasks and hence should be emphasized with higher weights when representing the query observations and the document models. Consequently, an initial attempt is made to compute the terms $c(w_i, Q)$ of a query word w_i and $P(w_j \mid D)$ of a document word w_j expressed in Eq. (7) for WTM, by taking into consideration the corresponding part-of-speech (POS) weights of the words (e.g., proper nouns: 1.5; nouns and verbs: 1.0; adjectives, adverbs and connectives: 0.6). Such an operation to some extent can be viewed as a kind of query/document expansion or reformulation [7]. The corresponding retrieval results of this preliminary investigation are shown in Table 5, which indeed confirm our expectation that further incorporation of POS weighting seems to be beneficial for WTM in most conditions. Further investigation of using extra cues such as recognition confidence and prosody information for WTM modeling is currently undertaken.

5. CONCLUSIONS

In this paper, we have examined a word topic model (WTM) approach for SDR. The SDR experimental results seem to reveal that WTM is a promising alternative to the other retrieval models compared in this paper. WTM has also been applied to speech recognition and summarization [13, 14], demonstrated with good potential. Future work on WTM includes discriminative model training [4] and integration with the other more sophisticated indexing mechanisms [6, 15] for larger-scale SDR tasks.

6. ACKNOWLEDGEMENTS

This work was supported in part by the National Science Council, Taiwan, under Grants: NSC96-2628-E-003-015-MY3, NSC95-2221-E-003-014-MY3, and NSC97-2631-S-003-003.

7. REFERENCES

[1] J. M. Ponte and W. B. Croft, "A language modeling approach to information retrieval", in *Proc. ACM SIGIR 1998*.
[2] D. Miller et al., "A hidden Markov model information retrieval System", in *Proc. ACM SIGIR 1999*.
[3] T. Hoffmann, "Unsupervised learning by probabilistic latent semantic analysis," *Machine Learning* 42, 2001.
[4] B. Chen et al. "A discriminative HMM/N-gram-based retrieval approach for Mandarin spoken documents," *ACM Trans. on Asian Language Information Processing* 3(2), 2004.
[5] L. S. Lee and B. Chen, "Spoken document understanding and organization," *IEEE Signal Processing Magazine* 22(5), 2005.
[6] C. Chelba et al., "Retrieval and browsing of spoken content," *IEEE Signal Processing Magazine* 25(3), 2008.
[7] C. D. Manning ta al., *Introduction to Information Retrieval*, Cambridge University Press, 2008.
[8] T. Joachims and F. Radlinski, "Search engines that learn from implicit feedback," *IEEE Trans. Computer* 40(8), 2007.
[9] D. M. Blei et al., "Latent Dirichlet allocation," *Journal of Machine Learning Research* 3, 2003.
[10] L. Saul and F. Pereira, "Aggregate and mixed-order Markov models for statistical language processing," in *Proc. EMNLP 1997*.
[11] R. Nallapati, "Discriminative models for information retrieval," in *Proc. ACM SIGIR 2004*.
[12] R. Nallapati, "Discriminative models for information retrieval," in *Proc. ACM SIGIR 2004*.
[13] B. Chen, "Word topic models for spoken document retrieval and transcription," *ACM Transactions on Asian Language Information Processing*, March 2009.
[14] Y.T. Chen, et al., "A probabilistic generative framework for extractive broadcast news speech summarization," *IEEE Trans. Audio, Speech and Language Processing* 17(1), 2009.
[15] T. Chia et al., "A lattice-based approach to query-by-example spoken document retrieval," in *Proc. ACM SIGIR 2008*.